



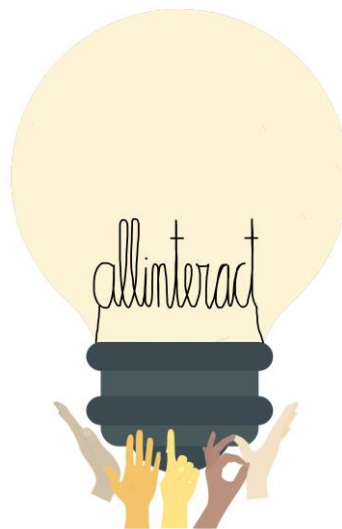
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement num. 872396



ALLINTERACT

SOCIAL MEDIA ANALYTICS PROTOCOL

February 2021



Allinteract – Social Media Analytics Protocol by Flecha, R. & Pulido, C is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

How to cite this item:

Flecha, R. Pulido C. (2021). Allinteract-Social Media Analytics Protocol is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License is available in https://archive.org/details/@crea_research



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement num. 872396



Project Number: 872396

Project Acronym: ALLINTERACT

Project title: Widening and diversifying citizen engagement in science

SMA protocol

This document compiles some recommendations to be considered for the elaboration of the SMA protocol.

SOCIAL MEDIA ANALYTICS

1) Selection of social networks, pages and hashtags

According to the Grant Agreement, Social Media Analytic will aim to identify and provide information on which awareness-raising actions and policies that succeed at linking scientific advancements that citizens appreciate to the research that led to them and foster, in turn, their participation in scientific research with social impact and the recruitment in science. To fulfil this purpose, SMA will seek to reach citizens' voices, paying special attention on vulnerable groups and young citizens. For this reason, different social networks will be included in this analysis. The following table 1 compiles the main characteristics of each social network that need be taken into account before the SMA.

Table 1. Characteristics of Social Media networks

Network	Options		Characteristics
Facebook	Pages		Most of Facebook pages are not individual initiatives, but they belong to organizations, associations, media or campaigns. Citizens like or follow these pages, so they can see in their home page the contents shared by these pages. When pages share their content, citizens interact with this content through likes, sharing or comments, so this is one way to access citizens' voices
	Groups	Public	Public groups include the voices of citizens, as all those users who are members of the group can post in the group and interact with the shared content. Users who are not part of the group can find the group (as long as the group is visible) and can see the content and interactions shared within the group, but they can only interact with the content through likes (they cannot publish a new post or comment in existing posts)
		Private	Allinteract will not access private groups
	Individual profiles		One option to search what individual citizens post in their timelines is to browse by hashtags (e.g. #gender). The results will show only those posts containing the hashtag, which are public, including pages and those citizens with a public Facebook profile. NOTE: It is important to ensure that the results obtained through this search are representative of society, because the use of hashtag is not broadly normalized yet and most users do not post using hashtags. In fact, the use of



		<p>hashtags is more common in pages or institutional profiles than in individual profiles.</p> <p>NOTE: Most citizens do not have their profiles in a public option, so it is difficult to get the whole picture in a search using hashtags.</p> <p>The second option is to use the keyword instead of the hashtag (without the #) in the browser and refining by “publications”. However, the results in this search show only publications containing the keyword posted by your friends or public pages. Therefore, this is not a real alternative, as it is impossible to see the whole picture in these results.</p>
Twitter	Individual profiles	The search can be carried out through hashtags and the results will show all public posts posted by individuals using the hashtag. Interactions with other citizens can be analysed through RT, comments or likes
Instagram	Individual profiles	The search can be carried out through hashtags and the results will show all public posts posted by individuals using the hashtag. Interactions with other citizens can be analysed mainly through likes. This social network is the most visual of all and users share pictures with hashtags and sometimes, a brief description. Therefore, it contains less written interactions among users.
Reddit	SubReddit	<p>When you conduct a search in Reddit, you can obtain results by posts and communities/users. As reddit does not use #, the posts results show all posts that use the keyword in every public reddit.</p> <p>When the search is refined by communities, you obtained all communities that mention that word (sorted by relevance, new... and filtered by period of time). In the communities, users share posts and comment posts about different topics. Some communities have categories to filter the posts (e.g: by topic, type of post, location...)</p>
	Individual users	Individual users can join Reddit communities and post or comment existing posts

ALLINTERACT Description of Action (p. 9) details that the selection of data from social media will follow a twofold approach. First of all, it is important to clarify bot approaches as defined by Cabré et al. (2017)

- a) **Top – down strategy:** this approach consists in the definition of keywords related to gender/education and the topics of the project to contrast if these topics are expressed by citizens in their opinions, comments and post expressed in social media.
- b) **Bottom-up strategy:** this approach consists in the identification of topics that emerge from those keywords and hashtags most used by citizens in twitter. Then the emerging topics are contrasted with the topics defined by the project and analyse if they cover all aspects proposed by the project and/or there are some additional issues that were not covered initially by the project.

The SMA protocol is drawn upon a pilot conducted during December – January. The main conclusions of the pilot and the results of the testing were essential to nourish this protocol and provide specific guidelines.

In both strategies (top-down and bottom-up) there are two teams, one focused on gender and the other in education.

a) **Top-down**

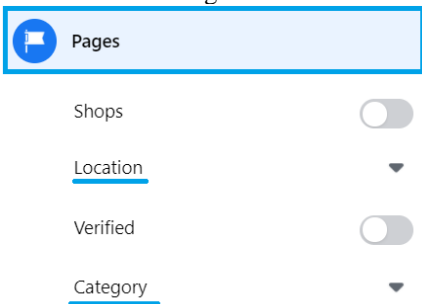
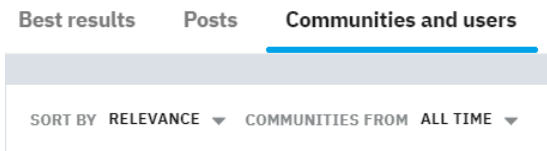
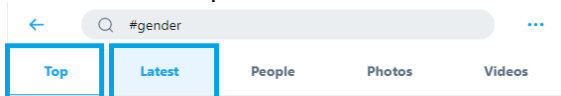
This strategy will include the analysis of 4 social networks: Twitter, Facebook, Instagram

Taking into account the main characteristics of each social network, the selection of 5 top pages or hashtags requires the implementation of different searching procedures. In all cases, keywords and

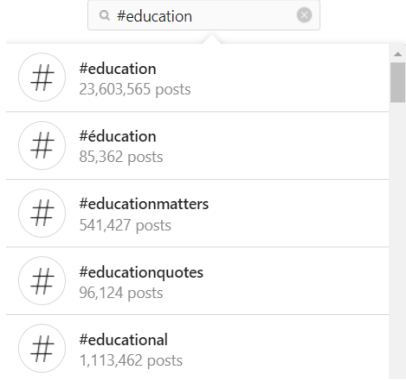


hashtags **will be in English** in order to obtain a global insight of interactions of European citizens in social media. The following table 2 summarizes all the procedures as well as the criteria for the selection of Facebook pages, subReddits and hashtags:

Table 2. Criteria for the selection of pages and hashtags.

Facebook	
Procedure of search	Criteria for the selection of pages
<ol style="list-style-type: none"> Combine two keywords in the browser: “gender” AND “women, equality, empowerment, feminism, rights, awareness...” Refine by “pages” and include all locations and categories 	<ul style="list-style-type: none"> High level of activity in the page (frequency and number of posts shared) High level of interaction with citizens (including likes and sharing, but mainly through comments) Outreach of the page (number of followers) Type of content shared in their posts (e.g. generation of debate, presence of scientific evidence)
Reddit	
Procedure of search	Criteria for the selection of communities
<ol style="list-style-type: none"> Search keyword (without hashtag) in the browser Refine by “Communities and Users” 	<ul style="list-style-type: none"> Interaction through debate (including dialogues and discussion among citizens on different topics and excluding humour pages) Outreach of the community (number of members) High level of interaction with users (including likes, posts and comments) Presence of filter categories that include topics related to the project (e.g. science, policies, news)
Twitter	
Procedure of search	Criteria for the selection of hashtags
<ol style="list-style-type: none"> Search trending hashtags on gender on https://ritetag.com/best-hashtags-for/gender Search each proposed hashtag on twitter using the symbol # Include “Top” and “latest” tweets 	<ul style="list-style-type: none"> Presence of citizens in the tweets (more individuals than institutions) Popularity of the hashtag (use of the hashtag and number of tweets) Correspondence between the content of the tweet and project topics High level of interactions among citizens (including likes, but mainly through RT and comments)
Instagram	
Procedure of search	Criteria for the selection of hashtags



<ol style="list-style-type: none"> 1. Search a keywords in the browser using the symbol # (note: chose general keywords that may be more used and contained in other hashtags) 2. Select most popular hashtags containing the keyword 3. Search all posts using the selected hashtags, including top and most recent posts  <p>The screenshot shows a search bar with '#education' entered. Below the search bar, a list of suggested hashtags is displayed, each with a circular icon containing a hash symbol and the number of posts associated with it:</p> <ul style="list-style-type: none"> #education: 23,603,565 posts #éducation: 85,362 posts #educationmatters: 541,427 posts #educationquotes: 96,124 posts #educational: 1,113,462 posts 	<ul style="list-style-type: none"> • Popularity of the hashtag (use of the hashtag and number of posts) • Correspondence between the content of the post and project topics • High level of interactions among citizens (including likes and comments)
---	---

- The extraction includes a maximum of 10,000 tweets/messages from the week before the extraction, under each hashtag or page.

b) **Bottom – up strategy:**

This methodology is only conducted in Twitter for technical limitations. During a week, each day, Top 50 Trending Topic in all EU countries are retrieved in order to identify hashtags related to gender and education. Moreover, additional strategies in education and gender are conducted:

- **Education:** this strategy relies on Twitter topics, so as to gather citizens' voices and perspectives on education. In this line, topics related to education will be tracked including, for example:
 - Education
 - College life
 - Graduate School
 - Home schooling
 - Language Learning
 - Online Education
- **Gender:** This strategy on gender will be implemented through the identification and compilation of messages under trending topics in gender. Additionally, trending topics in some relevant dates, such as International Women's Day or International Day of Women in Science are tracked. In order to identify and obtain trending topics in gender, specific websites can be used, such as:
 - **Trends 24** (<https://trends24.in/>): For the last 24 hours, compile the 10 Trending Topic worldwide for each hour and, among the trending topic, select those related to gender or education
 - **Get Day Trends** (<https://getdaytrends.com/>): For the last 24 hours, compile the 50 Trending Topic worldwide for each hour and, among the trending topic, select those related to gender or education



In both topics, gender and education, the top 10.000 messages by number of interactions under each topic (education) or Trending Topic (gender) are selected for extraction. Messages are retrieved using PYTHON and NVIVO, as detailed in section 5 of this protocol

2) Definition of the period for data collection

As mentioned above, UB will undertake the data collection following top-down and bottom-up strategies, which will need different periods for data collection:

- The extraction of messages and tweets will include those messages from Top-Down strategy as well as Bottom-up strategy.
- The selected messages are extracted using PYTHON and NVIVO and exported to an excel file, as detailed in section 5.
- All messages are ranked using criteria of most interactions (e.g. number of likes, RT and sharings)
- Top 10,000 messages by number of RT per hashtag are included in the analysis.

3) Definition of units of analysis

In order to unify the extracted data, it is important to establish the criteria that will rule the further analysis of the information. In this regard, some suggestions that could be taken into account are;

- To consider the complete tweet/post as the unit of analysis
- To include in the unit of analysis if the tweet/post has a link providing information
- To include the number of RT/sharing and comments obtained.
- To check the profile of the user to ensure that it is a true profile and not a fake one or a bot
- To ensure the final units of analysis meet the terms of service established in each social network
- The unit of analysis will also include possible images, videos and links accompanying the message
- In Instagram, where the most important part of the message is the image, researchers will analyze the content of the image and the message that this image transmits.
- All partners will need to access, check and consider the whole unit of analysis as part of the codification process.
- In those cases of multiple levels of hyperlinks, at least 2 levels (i.e. a link inside another link contained in the tweet/message), will be accessed, checked and considered.

4) Compliance with Terms and Conditions of Social Networks

- The Terms and Conditions of each social network have been carefully read in order to ensure that ALLINTERACT SMA complies with all these conditions.
- Special attention will be paid to data policy, how the consortium will access and use the information
- ALLINTERACT consortium is totally against all mechanisms that go against these terms and conditions, such as the creation of fake accounts to access private information
- Some websites that may be useful when ensuring the compliance with terms and conditions are: <https://research.fb.com/data/>, Public API (<https://developer.twitter.com/en/docs>), twitter for Websites (<https://developer.twitter.com/docs/twitter-for-websites/overview>), Twitter Cards (<https://developer.twitter.com/docs/tweets/optimize-with-cards/guides/getting-started>), or Developer Agreement (<https://developer.twitter.com/en/developer-terms/agreement>), Developer Policy (<https://developer.twitter.com/en/developer-terms/policy>) and Twitter Terms and Conditions <https://developer.twitter.com/en/developer-terms> .

5) Extraction of selected messages



- The extraction of the data under each page or hashtag is carried out using PYTHON and NVIVO. PYTHON programming language is a software promoted by Python Software Foundation (PSF) that enables the extraction of messages and information from social media.
- Information retrieved will include not only the message, but also message ID, time, number of likes and number of RT to identify those messages with more interactions among users.
- The extraction of data will respect General Data Protection Regulation (EU) 2016/679 as well as Terms and Conditions of the selected social networks. Collected data will include only public data that users consented to share, and all data will be rigorously anonymized using Amnesia, as detailed in the Data Management Plan and in Document 4.8.

6) Criteria of inclusion

- To be shared or published in the established period of time
- To be shared by a real user (not by a fake account or bot)
- To include the selected hashtags or keywords or to be posted in the selected pages
- To be among the 10,000 messages with more number of interactions among users

7) Exclusion criteria

- Not meeting the selection criteria
- Not meeting the scope of the topics (a-e) although containing the hashtags in gender/education

8) Anonymization

- All data is anonymized in order to follow the GDPR and Terms and Conditions of all social networks.
- Researchers will check that the selected messages meet the inclusion criteria and discard the others.
- They will also discard duplicated messages and messages posted by fake accounts or bots. For this purpose, researchers can use the following online tools:
 - o <https://botometer.osome.iu.edu/>
 - o <https://igauditor.com/>
 - o <https://influencermarketinghub.com/instagram-fake-follower-bot-checker-free/> (if an account has mainly bot followers, it will also be a bot)
 - o <https://www.sotrender.com/blog/2018/10/quick-guide-identifying-bots-and-trolls/>
- All messages that meet the inclusion criteria will be compiled in an excel file.
- Messages will be ranked by the number of interactions (nº of RT and sharing) and top 10,000 messages for each keyword or hashtags will be selected for analysis.
- Gathered data will comply with the EU GDPR and with the Terms and Conditions of each company

9) Codification of the data

- The codification process is conducted following the codebook, detailed below (table 3). This process includes two steps:
 - o Assignment of codes 1-5 according to **topics** a-e.
 - a. How citizens' benefit from scientific research
 - b. Citizen awareness of the impact of scientific research
 - c. Awareness-raising initiatives succeeding at engaging citizens in scientific



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement num. 872396







- participation, including the Open Access movement
- d. Awareness-raising actions that foster the recruitment of new talent in sciences
- e. Policies that promote awareness-raising actions and citizen engagement in science.
- Categorization of **scientific sources**, depending on the use or not of scientific evidence
- The codebook has been elaborated following a dialogic approach and it is detailed in table 3



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement num. 872396








Table 3. Dialogic Codebook

Code	Topic	Definition	Topic	
			Gender	Education
1	a. How citizens' benefit from scientific research	It includes those messages where citizens express how they implement or benefit from scientific research, although they are not aware that it is due to/do not mention scientific research	<p>Women neuropsychologists are integral to developing our understanding of the brain and #COVID19. For #InternationalDayOfWomen AndGirlsInScience, we spoke to @BPSDONChair to learn more about women's contribution in the field so far... @DONBPS https://t.co/iCRhECDBjz</p> 	<p>"When I became a teacher, I wanted to let each of my students know: 'I believe in you. You can do this hard thing.'" How teacher Denisha Saucedo finds the balance between showing students compassion and holding them to high expectations: link</p>  <p>Example of a kid from an educational evidence-based program</p>
2	b. Citizen awareness of the impact of scientific research	It includes both, initiatives that make citizens aware of the link between the benefits they appreciate and the research which led to them and also messages where citizens express their awareness of the impact of scientific research in education and gender.	<p>Data shows that girl children in India face immense discrimination while growing up. On the occasion of National Girl Child Day, let us promise to say no to gender inequality so that every girl child can #LiveTheJoy. #girls #Children #child #gender #growing #empower https://t.co/IKdR0nomf3</p> 	<p>Teaching is allowing others a "head start" in their long journey.</p> 



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement num. 872396



3	c. Awareness-raising initiatives succeeding at engaging citizens in scientific participation, including the Open Access movement	It includes those initiatives that enhance citizen participation in science	Calling all youth to join tomorrow to learn your duties and rights and build a gender equal world! Zoom Link :- https://t.co/ew22MbN0Fy @YouthCommunity #vartaleap #GenNationBuilding #Youth #gender #equality #rights https://t.co/tUKWklcLN		The joint programme & projects envisage the use of public #diplomacy in #science, #education, #culture, #sports, mass #media through the exchange of #experiences, #knowledge, & #ideas... https://t.co/VTwI7MoEqb	
4	d. Awareness-raising actions that foster the recruitment of new talent in sciences	It includes formal actions (e.g. campaign, workshops, seminars, conference, scholarships or vacancies specifically encouraging women, educational materials, interviews) and individual actions (e.g. personal experience or testimony)	It's #InternationalDayOf WomenAndGirlsIn Science& also #BlackHistoryMonth in the US, so we're highlighting some Black women scientists whose contributions we're all lucky to have! Who would you add to the list? https://t.co/417xRSFB1R		This "Periodic Table of Black History" celebrates Black writers, scientists, artists, politicians and more – from throughout history up to the present day! What (virtual or in-person) displays have you loved recently? (via educator Mr. Bedford)	
5	e. Policies that promote awareness-raising actions and citizen engagement in science.	It includes messages that mention policies, programs or institutional measures and initiatives that promote citizen engagement in science, representation of women scientist, or the creation of spaces for the participation of vulnerable groups in science	The local government in #Shenzhen has launched an official guidebook to identify and prevent #gender #discrimination in #advertisement, following a huge controversy caused by PurCotton, a brand based in the city. https://t.co/lw70c4XNpT		Teaching is allowing others a "head start" in their long journey.	
0	Not valid	It includes those messages that are no related to project topics	I hope it's a boy #Gender-reveal		"Fill in the blank! To me, ""teaching"" means:....."	



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement num. 872396



"blank"	Empty or not accessible	It includes those messages that cannot be accessed due to a broken link or containing a message that could not be extracted (e.g. emojis...)	
Scientific evidence			
Code	Scientific evidence	Definition	Example
0	Not Scientific Evidence	The message does not contain scientific evidence	"This ""Periodic Table of Black History"" celebrates Black writers, scientists, artists, politicians and more – from throughout history up to the present day! What (virtual or in-person) displays have you loved recently? (via educator Mr. Bedford)"
1	Certified Scientific Evidence	The message includes scientific evidence (articles from journals indexed in Scopus o JCR)	Recent research suggests that the more social pressure a man feels to be masculine, the more aggressive he may be - @newsmedical https://news-medical.net/news/20210128/Mens-aggression-is-tied-to-social-pressure-shows-study.aspx #gender
2	Supposed Scientific Evidence	The message says that it is based on scientific evidence, but there is no reference to the article or study	Yes!!!! And some Scientists believe we need Diversity in our Genes to survive long term!!!

Emergent categories

During the codification, a new category emerged because researchers found that citizens used Social Media to claim scientific evidence to prove users' claims and also used scientific arguments to discuss the validity of the sources provided. Therefore, a new code 3 is added to the codification of scientific evidence (Table 4)

Table 4. Emergent Category.

Code	Scientific evidence	Definition	Example
3	Reclaim Scientific Evidence	The user reclaims a scientific evidence or source to prove someone's argument or uses scientific claims to discuss the validity of the sources provided	Not exactly current though, and I don't trust any research about marijuana during the high point of the war on drugs. Are there any more recent studies?



- During the codification, first one researcher reads and analyse the whole unit of analysis, and assign to the message one number 0-5 for project topic (a-e) and another number 0-2 according to the sources referred.
- Then a double check process is carried out by a different researcher.
- When researchers find a message that cannot be clearly classified, this message is dialogically discussed, and researchers reach an agreement basing their decision on validity claims
- If the post/tweet is written in a language that it is not English, researchers translate the content in order to be able to codify the message.

10) Analysis of the data

- Once the dataset for each topic is compiled, the messages within each dataset can be analysed.
- For the analysis, the information of the message is used to fill those columns of the grid that apply. If there is not enough information in the selected messages to fill all the columns, columns can be left blank.
- A communicative analysis will be conducted.
 - *Transformative dimension*: will include those messages that evidence dimension what facilitates the social/political/scientific impact mentioned in the message or what facilitates the implementation of the targeted actions or policies
 - *Exclusionary dimension*: will include those messages that show obstacles hindering the achievement of the targeted social/political/scientific impact or the implementation of the targeted actions/policies
- The following table 5 is used to inform the analysis in each topic.

Table 5. Table of analysis under each topic

TOPIC a) How citizens' benefit from scientific research							
	<i>Benefits from social impact (i.e. direct benefits for social actors)</i>		<i>Benefits from political impact (i.e. evidence-based policies which have led to benefits for social actors);</i>		<i>Benefits from scientific impact (i.e. scientific research which have contributed to the advancement of science with social impact).</i>		
Transformative Dimension							
Exclusionary Dimension							
TOPIC b) Citizen awareness of the impact of scientific research							
	<i>Aim of the awareness-raising initiative</i>	<i>Promoting organization (i.e. aim; private/ public)</i>	<i>Level of intervention (i.e. local/national/ international; bottom-up/top-down)</i>	<i>Field of intervention</i>	<i>Audience (targeted and real)</i>	<i>Role of the targeted audience</i>	<i>Level of access to scientific evidence</i>



Transformative Dimension								
Exclusionary Dimension								
TOPIC c) Awareness-raising initiatives succeeding at engaging citizens in scientific participation, including the Open Access movement								
	<i>Aim of the enhancing/ participation action</i>	<i>Promoting organization (i.e. aim; private/ public)</i>	<i>Level of intervention (i.e. local /national/ international; bottom-up/ top- down)</i>	<i>Field of intervention</i>	<i>Participants (targeted and real)</i>	<i>Type of participation (role of the participants)</i>	<i>Level of access to scientific evidence</i>	<i>Social Impact</i>
Transformative Dimension								
Exclusionary Dimension								
TOPIC d) Awareness-raising actions that foster the recruitment of new talent in sciences								
	<i>Aim of the action targeting new talent for science</i>	<i>Promoting organization (i.e. aim; private/ public)</i>	<i>Level of intervention (i.e. local /national/ international; bottom-up/ top-down)</i>	<i>Field of intervention</i>	<i>Participants (targeted and real)</i>	<i>Type of participation (role of the participants)</i>	<i>Level of access to scientific evidence</i>	<i>Social Impact</i>
Transformative Dimension								
Exclusionary Dimension								
TOPIC e) Policies that promote awareness-raising actions and citizen engagement in science.								
	<i>Aim of the enhancing/ participation</i>	<i>Promoting organization (i.e. aim;</i>	<i>Level of intervention (i.e. local</i>	<i>Field of intervention</i>	<i>Participants (targeted and</i>	<i>Type of participation (role of the</i>	<i>Level of access to scientific</i>	<i>Social Impact</i>



	<i>action</i>	<i>private/ public)</i>	<i>/national/ international; bottom-up/ top-down)</i>		<i>real)</i>	<i>participants)</i>	<i>evidence</i>	
Transformative Dimension								
Exclusionary Dimension								

- During this analysis, new inductive categories may emerge within each topic a-e
- These new inductive categories are defined through consensus among different researchers. During this process, researchers propose and share the emerging category to the other researchers in a dialogic forum.
- After all messages of the dataset are analyzed, quantitative and qualitative evidence are elaborated, calculating the number of tweets in each category, and analyzing each message's content (transformative and exclusionary dimensions)
- The assessment of the reliability of data coding will also follow a dialogical approach, which needs "time, plurality and egalitarian dialogue, but this time use is efficient in order to guarantee the reliability of the analysis obtained. Dialogue is present throughout the process to guarantee the consistency of the analysis" (Pulido et al., 2020, p. 384).



SOCIAL MEDIA COMMUNICATIVE OBERVATION

1. Selection of Facebook groups and subreddits

- Following the procedure detailed below, a search of the most popular public groups in gender and education in Facebook and public Reddit communities is conducted for data collection.
- The search of groups and communities follows the following process in table 6.

Table 6. Process to select groups for the Social Media Communicative Observation.

Facebook	Reddit
<ul style="list-style-type: none"> ○ Combine two keywords in the browser: “gender” AND “women, equality, empowerment, feminism, rights, awareness...” ○ Refine by “groups” and include only public groups 	<ol style="list-style-type: none"> 1. Search keyword (without hashtag) in the browser 2. Refine by “Communities and Users” and include communities from all time

- Criteria for the selection of Facebook public groups and subreddits
 - Inclusion of young people,
 - Inclusion of vulnerable groups
 - Representation of diverse citizens in terms of age, gender, academic level or cultural background, among others.
 - Representation of different countries
 - Number of users/members
- At least **two groups** in each social media will target young citizens and vulnerable groups.

2. Obtention of informed consent

- The Chair of the Ethics & Data Management Committee contacts the administrators of these groups and will explain them the objectives of the project and of the Social Media Communicative Observation.
- If a group administrator does not answer within **5 days** or does not agree to participate, the next group of the list will be selected and contacted
- The group administrators inform the members and ask for their consent. ALLINTERACT researchers provide consent forms to group members who agree to participate in Social Media Communicative Observation prior to the data collection.

3. Introduction of scientific evidence

- Evidence of the social impact of scientific results on gender and educational research will be introduced in each group every week.
- The list of evidence of scientific improvements in gender and education will be obtained from the results of the Literature Review.
- In order to foster citizens’ participation, researchers ensure all scientific evidence is understandable and written in a non-specialized language. The discussion is shared in the



natural language of users.

- At least one statement of scientific evidence per week will be shared in each group, but if group members and the administrator agree, up to 3 statements per week may be shared.

4. Communicative observation of the interactions among users

- Researchers will interact with users and enter in discussion with them about the scientific evidence shared in the group.
- Researchers will monitor how users' interactions change after having introduced scientific data (e.g. use of scientific arguments in the debate) as well as the correlation between evidence mentioning and participation in scientific research (e.g. amount of users participating in the debate).

5. Extraction of data

- The extraction of data is carried out using PYTHON and NVIVO.
- Only interactions among users who have provided their consent are monitored and analysed.

6. Anonymization of data

- All data is anonymized all data in order to follow the GDPR and Terms and Conditions of each social network (Facebook and REDDIT).

7. Analysis

- The analysis combines predefined categories with those categories that emerged during the process.
- The initial categories emerge from project topics (a-e), each partner being in charge of the same topic as in the literature review:
 - How citizens' **benefit** from scientific research
 - Citizen **awareness** of the impact of scientific research
 - Awareness-raising initiatives succeeding at **engaging** citizens in scientific participation, including the Open Access movement
 - Awareness-raising actions that foster the **recruitment** of new talent in sciences
 - Policies** that promote awareness-raising actions and citizen engagement in science.
- A communicative analysis is be conducted. The following table 7 can be used to inform the initial analysis

Table 7. Table of Analysis for the Social Media Communicative Observation

TOPIC a) How citizens' benefit from scientific research			
	Benefits from social impact (i.e. direct benefits for social actors)	Benefits from political impact (i.e. evidence-based policies which have led to benefits for social actors);	Benefits from scientific impact (i.e. scientific research which have contributed to the advancement of science with social impact).



Transformative Dimension								
Exclusionary Dimension								
TOPIC b) Citizen awareness of the impact of scientific research								
	<i>Aim of the awareness-raising initiative</i>	<i>Promoting organization (i.e. aim; private/ public)</i>	<i>Level of intervention (i.e. local/national/ international; bottom-up/top-down)</i>	<i>Field of intervention</i>	<i>Audience (targeted and real)</i>	<i>Role of the targeted audience</i>	<i>Level of access to scientific evidence</i>	
Transformative Dimension								
Exclusionary Dimension								
TOPIC c) Awareness-raising initiatives succeeding at engaging citizens in scientific participation, including the Open Access movement								
	<i>Aim of the enhancing/ participation action</i>	<i>Promoting organization (i.e. aim; private/ public)</i>	<i>Level of intervention (i.e. local /national/ international; bottom-up/ top-down)</i>	<i>Field of intervention</i>	<i>Participants (targeted and real)</i>	<i>Type of participation (role of the participants)</i>	<i>Level of access to scientific evidence</i>	<i>Social Impact</i>
Transformative Dimension								
Exclusionary Dimension								
TOPIC d) Awareness-raising actions that foster the recruitment of new talent in sciences								
	<i>Aim of the action targeting new talent for science</i>	<i>Promoting organization (i.e. aim; private/</i>	<i>Level of intervention (i.e. local /national/</i>	<i>Field of intervention</i>	<i>Participants (targeted and real)</i>	<i>Type of participation (role of the</i>	<i>Level of access to scientific</i>	<i>Social Impact</i>



		<i>public)</i>	<i>international; bottom-up/ top-down)</i>			<i>participants)</i>	<i>evidence</i>	
Transformative Dimension								
Exclusionary Dimension								

TOPIC e) Policies that promote awareness-raising actions and citizen engagement in science.

	<i>Aim of the enhancing/ participation action</i>	<i>Promoting organization (i.e. aim; private/ public)</i>	<i>Level of intervention (i.e. local /national/ international; bottom-up/ top-down)</i>	<i>Field of intervention</i>	<i>Participants (targeted and real)</i>	<i>Type of participation (role of the participants)</i>	<i>Level of access to scientific evidence</i>	<i>Social Impact</i>
Transformative Dimension								
Exclusionary Dimension								

- New inductive categories may emerge from the analysis and they are defined through consensus among different researchers. During this process, researcher propose and share the emerging category to other researchers through a message in the workspace forum.
- Researchers from each institution will analyse each message collaboratively and dialogically, building agreements based on validity claims
- An analysis of the correlation is conducted between awareness of the social impact of research and engagement in science on the previous selection of individuals. This analysis will take as variables “citizen awareness of the social impact of research” and “citizen engagement in science”. The correlation between both variables will be crossed using R. Since both variables are nominal and dichotomic, Phi, Cramer’s V and the Contingency coefficient will be selected. To observe directionality, Lambda, Goodman and Kruskal’s Tau and the Uncertainty coefficient will be used.